

# A Note on Matrix Concentration Inequalities via the Method of Exchangeable Pairs

Daniel Paulin

*Department of Mathematics, National University of Singapore  
10 Lower Kent Ridge Road, Singapore 119076, Republic of Singapore.  
e-mail: [paulindani@gmail.com](mailto:paulindani@gmail.com)*

**Abstract:** The aim of this paper is to prove an improved version of the bounded differences inequality for matrix valued functions (see [Tropp \(2012\)](#), Corollary 7.5), by developing the methods of [Mackey et al. \(2012\)](#). Along the way, we prove new trace inequalities for the matrix exponential.

**AMS 2000 subject classifications:** Primary 60E15, 60B20; secondary 15A42, 15A45.

**Keywords and phrases:** Concentration inequality, random matrices, exchangeable pairs, Stein's method, trace inequality.

## 1. Introduction

Let  $Z_1, \dots, Z_n$  be independent (or dependent) random variables, and  $\mathbf{X} = f(Z_1, \dots, Z_n)$  be a random Hermitian matrix. One specific example is when  $\mathbf{X} = \sum_k \mathbf{X}_k$  is a sum of random matrices. In many situations, we are interested in bounding the quantity  $\mathbb{P}(\lambda_{\max}(\mathbf{X}) \geq t)$ .

[Ahlsweede and Winter \(2002\)](#) was the first to use Laplace transform method in this setting, they show that for any random Hermitian matrix  $\mathbf{X}$ ,

$$\mathbb{P}(\lambda_{\max}(\mathbf{X}) \geq t) \leq \inf_{\theta > 0} \left\{ e^{-\theta t} \mathbb{E} \operatorname{tr} \exp(\theta \mathbf{X}) \right\}, \quad (1.1)$$

thus for  $\mathbf{X} = \sum_k \mathbf{X}_k$ ,

$$\mathbb{P}(\lambda_{\max}(\mathbf{X}) \geq t) \leq \inf_{\theta > 0} \left\{ e^{-\theta t} \mathbb{E} \operatorname{tr} \exp \left( \theta \sum_k \mathbf{X}_k \right) \right\}. \quad (1.2)$$

Estimating the right hand side now poses a difficulty, because in general,  $e^{A+B} \neq e^A \cdot e^B$  for the matrix exponential.

[Tropp \(2012\)](#) proves the following lemma to estimate the right hand side:

**Lemma 1.1** (Lemma 3.4 of [Tropp \(2012\)](#)). *Consider a finite sequence  $\{\mathbf{X}_k\}$  of independent, random, self-adjoint matrices. Then*

$$\mathbb{E} \operatorname{tr} \exp \left( \sum_k \theta \mathbf{X}_k \right) \leq \operatorname{tr} \exp \left( \sum_k \log \mathbb{E} e^{\theta \mathbf{X}_k} \right) \text{ for } \theta \in \mathbb{R}.$$

This Lemma is based on a corollary of [Tropp \(2012\)](#) (which is derived from a Theorem of Lieb):

**Corollary 1.1** (Corollary 3.3 of [Tropp \(2012\)](#)). *Let  $\mathbf{H}$  be a fixed self-adjoint matrix, and let  $\mathbf{X}$  be a random self-adjoint matrix. Then*

$$\mathbb{E} \operatorname{tr} \exp(\mathbf{H} + \mathbf{X}) \leq \operatorname{tr} \exp(\mathbf{H} + \log(\mathbb{E} e^{\mathbf{X}})).$$

These inequalities are used in [Tropp \(2012\)](#) to prove matrix versions of various concentration inequalities for sums of random matrices (Chernoff, Bernstein), and inequalities for matrix martingales (Azuma-Hoeffding, and matrix bounded differences).

[Mackey et al. \(2012\)](#) takes a different approach. They make the following basic definition ([Mackey et al. \(2012\)](#), Definition 2.2):

**Definition 1** (Matrix Stein Pair). *Let  $(Z, Z')$  be an exchangeable pair of random variables taking values in a polish space  $\mathcal{Z}$ , and let  $\Psi : \mathcal{Z} \rightarrow \mathbb{H}^d$  be a measurable function. Define the random Hermitian matrices*

$$\mathbf{X} := \Psi(Z) \quad \text{and} \quad \mathbf{X}' := \Psi(Z').$$

We say that  $(\mathbf{X}, \mathbf{X}')$  is a matrix Stein pair if there is a constant  $\alpha \in (0, 1]$  for which

$$\mathbb{E}(\mathbf{X} - \mathbf{X}' | Z) = \alpha \mathbf{X} \text{ almost surely.} \quad (1.3)$$

The constant  $\alpha$  is called the scale factor of the pair. When discussing a matrix Stein pair  $(\mathbf{X}, \mathbf{X}')$ , we always assume that  $\mathbb{E} \|\mathbf{X}\|^2 < \infty$ .

Suppose that  $(\mathbf{X}, \mathbf{X}')$  is a matrix Stein pair, then they write the derivate moment generating function of  $m(\mathbf{X})$  as

$$\begin{aligned} m(\theta)' &= \mathbb{E} \operatorname{tr}(\mathbf{X} e^{\theta \mathbf{X}}) = \mathbb{E} \operatorname{tr} \left( \frac{1}{\alpha} (\mathbf{X} - \mathbf{X}') e^{\theta \mathbf{X}} \right) \\ &= \mathbb{E} \operatorname{tr} \left( \frac{1}{2\alpha} (\mathbf{X} - \mathbf{X}') (e^{\theta \mathbf{X}} - e^{\theta \mathbf{X}'}) \right), \end{aligned}$$

using exchangeability in the last step.

To further bound this quantity, they prove the following trace inequality (Lemma 3.4 of [Mackey et al. \(2012\)](#)):

**Lemma 1.2.** *Let  $I$  be an interval of the real line. Suppose that  $g : I \rightarrow \mathbb{R}$  is a weakly increasing function and that  $h : I \rightarrow \mathbb{R}$  is a function whose derivative  $h'$  is convex. For all matrices  $\mathbf{A}, \mathbf{B} \in \mathbb{H}^d(I)$ ,*

$$\begin{aligned} &\operatorname{tr}[(g(\mathbf{A}) - g(\mathbf{B})) \cdot (h(\mathbf{A}) - h(\mathbf{B}))] \\ &\leq \frac{1}{2} \operatorname{tr}[(g(\mathbf{A}) - g(\mathbf{B}))(\mathbf{A} - \mathbf{B}) \cdot (h'(\mathbf{A}) + h'(\mathbf{B}))]. \end{aligned}$$

When  $h'$  is concave, the inequality is reversed. The same results hold for the standard trace.

This lemma is based on a standard trace inequality ([Petz \(1994\)](#), Proposition 3).

**Corollary 1.2.** *For  $\theta > 0$ ,*

$$\operatorname{tr} \left( (\mathbf{X} - \mathbf{X}') (e^{\theta \mathbf{X}} - e^{\theta \mathbf{X}'}) \right) \leq \frac{\theta}{2} \operatorname{tr} \left( (\mathbf{X} - \mathbf{X}')^2 (e^{\theta \mathbf{X}} + e^{\theta \mathbf{X}'}) \right)$$

Using this corollary, we can bound the derivate of the trace mgf:

$$\begin{aligned} m(\theta)' &\leq \mathbb{E} \operatorname{tr} \left( \frac{\theta}{2\alpha} (\mathbf{X} - \mathbf{X}')^2 (e^{\theta \mathbf{X}} + e^{\theta \mathbf{X}'}) \right) \\ &= \mathbb{E} \operatorname{tr} \left( \frac{\theta}{\alpha} (\mathbf{X} - \mathbf{X}')^2 e^{\theta \mathbf{X}} \right), \end{aligned}$$

and this quantity can be bounded in many situations.

The advantage of this approach compared to [Tropp \(2012\)](#) is that the constants are often better, and some dependent cases can be also treated. The disadvantage is that other than sums of random matrices, few other cases can be written as Stein pairs. This means that matrix martingales, and the method of bounded differences, are not possible to recover.

The purpose of this paper is to show that [Mackey et al. \(2012\)](#) can be improved to show the method of bounded differences for matrix valued functions. We are going to prove new trace inequalities, which generalize Corollary 1.2, and allow us to go beyond Stein pairs.

Our inequality also works for weakly dependent random variables. We quantify the dependence by a matrix:

**Definition** (Dobrushin's interdependence matrix). *Let  $X := (X_1, \dots, X_n)$  be a random vector taking values in  $\Lambda := (\Lambda_1, \dots, \Lambda_n)$ , with law  $\mu$ . Suppose that  $D := (d_{ij})_{1 \leq i, j \leq n}$  is an  $n \times n$  matrix with nonnegative entries and zeroes on the diagonal such that for any  $i$ , and any  $x, y \in \Lambda$ ,*

$$d_{TV}(\mu_i(\cdot | x_{-i}), \mu_i(\cdot | y_{-i})) \leq \sum_{j=1}^n d_{ij} \mathbb{1}[x_j \neq y_j]$$

where  $d_{TV}$  is the total variational distance. Then we say that  $D$  is a Dobrushin interdependence matrix for the random vector  $X$  (or equivalently random measure  $\mu$ ). Here  $x_{-i} := (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$  and  $\mu_i(\cdot | x_{-i})$  is the conditional distribution of  $X_i$  given  $X_{-i} = x_{-i}$ .

Concentration inequalities for real valued functions Hamming Lipschitz functions under the condition  $\|D\|_2 < 1$  have been proven in [Chatterjee \(2005\)](#), Chapter 4.

## 2. Results

The following result is a strengthening of Corollary 7.5 of [Tropp \(2012\)](#). We have exponent  $-t^2/\sigma^2$  instead of  $-t^2/8\sigma^2$  in the independent case, and our result also works under Dobrushin-type weak dependence.

**Theorem 2.1.** *Let  $\{Z_k : k = 1, \dots, n\}$  be an independent family of random variables, and let  $\mathbf{H}$  be a function that maps  $n$  variables to a self adjoint matrix of dimension  $d$ . Consider a sequence  $\{\mathbf{A}_k\}$  of fixed self-adjoint matrices that satisfy*

$$(\mathbf{H}(z_1, \dots, z_k, \dots, z_n) - \mathbf{H}(z_1, \dots, z'_k, \dots, z_n))^2 \leq \mathbf{A}_k^2, \quad (2.1)$$

where  $z_i$  and  $z'_i$  range over all possible values of  $Z_i$  for each index  $i$ . Compute the variance parameter

$$\sigma^2 := \left\| \sum_k \mathbf{A}_k^2 \right\|. \quad (2.2)$$

Then for all  $t \geq 0$ ,

$$\mathbb{P} \{ \lambda_{\max} (\mathbf{H}(Z) - \mathbb{E}\mathbf{H}(Z)) \geq t \} \leq d \cdot e^{-t^2/\sigma^2} \quad (2.3)$$

where  $Z = (Z_1, \dots, Z_n)$ .

Alternatively, suppose that  $\{Z_k : k = 1, \dots, n\}$  is a family of dependent random variables with Dobrushin interdependence matrix  $D$ . If  $D$  satisfies  $\max(\|D\|_1, \|D\|_\infty) < 1$ , then for every  $t \geq 0$ ,

$$\mathbb{P} \{ \lambda_{\max} (\mathbf{H}(Z) - \mathbb{E}\mathbf{H}(Z)) \geq t \} \leq d \cdot e^{-t^2/(c\sigma^2)}, \quad (2.4)$$

with

$$c := \frac{1/(1 - \|D\|_1) + 1/(1 - \|D\|_\infty)}{2}. \quad (2.5)$$

A simple corollary of this the following matrix Hoeffding bound (in the independent case, similar to Corollary 4.2 of [Mackey et al. \(2012\)](#)):

**Corollary 2.1.** *Let  $\{\mathbf{Y}_k : k = 1, \dots, n\}$  be an independent family of  $\mathbb{H}^d$  matrices, and let  $\mathbf{H}$  be a function that maps  $n$  variables to a self adjoint matrix of dimension  $d$ . Consider a sequence  $\{\mathbf{A}_k\}$  of fixed self-adjoint matrices,*

$$\mathbb{E}\mathbf{Y}_k = 0, \quad \mathbf{Y}_k^2 \preceq \mathbf{A}_k^2, \quad (2.6)$$

Define the variance parameter

$$\sigma^2 := \left\| \sum_k \mathbf{A}_k^2 \right\|.$$

Then for all  $t \geq 0$ ,

$$\mathbb{P} \left\{ \lambda_{\max} \left( \sum_k \mathbf{Y}_k \right) \geq t \right\} \leq d \cdot e^{-t^2/(4\sigma^2)}.$$

Alternatively, for  $\{\mathbf{Y}_k : k = 1, \dots, n\}$  weakly dependent with Dobrushin matrix  $D$  satisfying  $\max(\|D\|_1, \|D\|_\infty) < 1$ , we have

$$\mathbb{P} \left\{ \lambda_{\max} \left( \sum_k \mathbf{Y}_k \right) \geq t \right\} \leq d \cdot e^{-t^2/(4c\sigma^2)},$$

with  $c$  defined as in (2.5).

**Remark 2.1.** The 4 in the exponent comes from the fact that (2.1) is satisfied for  $2\mathbf{A}_k$ .

An important tool in the proof is the following trace inequality:

**Theorem 2.2.** *Let  $A, B, C$  be Hermitian matrices of equal size, then*

$$\mathrm{tr} (C(e^A - e^B)) \leq \mathrm{tr} \left( \frac{C^2 + (A - B)^2}{2} \left( \frac{e^A + e^B}{2} \right) \right).$$

**Corollary 2.2.** *Under the same conditions, for  $\theta > 0$ ,*

$$\mathrm{tr} (C(e^{\theta A} - e^{\theta B})) \leq \theta \mathrm{tr} \left( \frac{C^2 + (A - B)^2}{2} \left( \frac{e^{\theta A} + e^{\theta B}}{2} \right) \right),$$

and for  $\theta < 0$ ,

$$\mathrm{tr} (C(e^{\theta A} - e^{\theta B})) \geq \theta \mathrm{tr} \left( \frac{C^2 + (A - B)^2}{2} \left( \frac{e^{\theta A} + e^{\theta B}}{2} \right) \right).$$

*Proof.* Apply Theorem 2.2 to  $\theta A, \theta B, \theta C$ . □

We also prove this result:

**Theorem 2.3.** (*Matrix Hölder inequality*) Let  $A, B, C, D$  be Hermitian matrices with  $A$  and  $B$  positive semidefinite, and  $0 \leq p \leq 1$ , then we have

$$\mathrm{Re} \left( \mathrm{tr} (C A^p D B^{1-p} + C A^{1-p} D B^p) \right) \leq \mathrm{tr} \left( \frac{C^2 + D^2}{2} (A + B) \right). \quad (2.7)$$

### 3. Proof of the bounded differences inequality

For a random matrix  $\mathbf{X}$ , the normalized trace mgf is defined, similarly to Definition 3.2 of Mackey et al. (2012), as

$$m(\theta) := m_{\mathbf{X}}(\theta) = \mathbb{E} \overline{\mathrm{tr}} e^{\theta \mathbf{X}} = \frac{1}{d} \mathbb{E} \mathrm{tr} e^{\theta \mathbf{X}},$$

which may not exist for all values of  $\theta$ .

We are going to use Proposition 3.3 of Mackey et al. (2012)

**Proposition 3.1.** (*Matrix Laplace Transform Method*) Let  $\mathbf{X}$  be a random matrix with normalized trace mgf  $m(\theta) := \mathbb{E} \overline{\mathrm{tr}} e^{\theta \mathbf{X}}$ . For each  $t \in \mathbb{R}$ ,

$$\mathbb{P}\{\lambda_{\max}(\mathbf{X}) \geq t\} \leq d \cdot \inf_{\theta > 0} \exp\{-\theta t + \log m(\theta)\}, \text{ and} \quad (3.1)$$

$$\mathbb{P}\{\lambda_{\min}(\mathbf{X}) \leq t\} \leq d \cdot \inf_{\theta < 0} \exp\{-\theta t + \log m(\theta)\}. \quad (3.2)$$

*Proof of Theorem 2.1.* We follow the Markov chain approach Chatterjee (2005).

As shown in Chapter 4, an exchangeable pair  $(X, X')$  automatically defines a reversible Markov kernel  $P$  as

$$Pf(X) := \mathbb{E}(f(X')|X = x),$$

where  $f$  is any function with  $\mathbb{E}|f(X)| < \infty$ . Suppose that  $X$  takes values in a Polish space  $\Omega$ , then

**Lemma 3.1** (Lemma 4.1 of Chatterjee (2005)). Suppose that  $f : \Omega \rightarrow \mathbb{R}$  is a measurable function with  $\mathbb{E}f(X) = 0$ , and there is a finite constant  $L$  such that

$$\sum_{k=0}^{\infty} |P^k f(x) - P^k f(y)| \leq L \text{ for every } x \text{ and } y, \quad (3.3)$$

then

$$F(x, y) := \sum_{k=0}^{\infty} (P^k f(x) - P^k f(y)) \quad (3.4)$$

satisfies  $F(X, X') = -F(X', X)$  and  $\mathbb{E}(F(X, X')|X) = f(X)$ .

With a simple adaptation of the proof, the reader can verify that this Lemma also holds for matrix valued functions  $f : \Omega \rightarrow \mathbb{H}^d$ , with (3.3) replaced by

$$\sum_{k=0}^{\infty} \|P^k f(x) - P^k f(y)\| \leq L \text{ for every } x \text{ and } y. \quad (3.5)$$

We need to define property **P** as in Chatterjee (2005):

**Definition.** Let  $\{X(k)\}_{k \geq 0}$  and  $\{X'(k)\}_{k \geq 0}$  be two chains from the kernel defined by  $(X, X')$ , for arbitrary initial values  $x, y \in \Omega$ . We say that a coupling of these two chains satisfies property **P** if for every  $x, y \in \Omega$ , and every  $k$ , the marginal distribution of  $X(k)$  only depends on  $x$ , and the marginal distribution of  $X'(k)$  only depends on  $y$ .

We propose the following matrix version Lemma 4.2 of Chatterjee (2005) (the proof can be easily adapted):

**Lemma 3.2.** Suppose that a coupling of  $\{X(k)\}_{k \geq 0}$  and  $\{X'(k)\}_{k \geq 0}$  satisfies property **P**. Let  $f : \Omega \rightarrow \mathbb{H}^d$  be a function such that  $\mathbb{E}f(X) = 0$ . Suppose that there exists  $L < \infty$  such that for every  $x, y \in \Omega$ ,

$$\sum_{k=0}^{\infty} \|\mathbb{E}(f(X(k)) - f(X'(k)) | X(0) = x, X'(0) = y)\| \leq L. \quad (3.6)$$

Then, the function  $F$  defined as

$$F(x, y) := \sum_{k=0}^{\infty} \mathbb{E}(f(X(k)) - f(X'(k)) | X(0) = x, X'(0) = y) \quad (3.7)$$

satisfies  $F(X, X') = -F(X', X)$  and  $\mathbb{E}(F(X, X') | X) = f(X)$ .

First, we will prove the independent case:

*Proof of independent case.* Let  $X := (X_1, \dots, X_n)$  be a vector with independent components ( $X_i$ : component  $i$ ,  $\{X(k)\}_{k \geq 0}$ : Markov chain).

Let

$$X^{(r)} := (X_1^{(r)}, \dots, X_n^{(r)}) \quad (3.8)$$

be independent copies of  $X$ , for  $r \geq 0$ . Let  $I, I_1, \dots, I_k \dots$  be uniformly distributed indexes in  $[n]$ , independent of each other and of  $X$  and  $X^{(r)}$ . Define  $X'$  as

$$X'_i = X_i \text{ for } i \neq I \text{ and } X'_I = X_I^{(0)}.$$

Now we are ready to construct  $X(k)$  and  $X'(k)$ :

Suppose that  $X(0) = x$  and  $X'(0) = y$ , for  $x, y \in \Omega$ . For  $k \geq 1$ , define  $X(k)$  as

$$X_i(k) := X_i(k-1) \text{ for } i \neq I_k \text{ and } X_{I_k}(k) := X_{I_k}^{(k)}.$$

Similarly, for  $k \geq 1$ , define  $X'(k)$  as

$$X'_i(k) := X'_i(k-1) \text{ for } i \neq I_k \text{ and } X'_{I_k}(k) := X_{I_k}(k).$$

With this definition,  $\{X(k)\}_{k \geq 0}$  and  $\{X'(k)\}_{k \geq 0}$  are having the same distribution as the Markov chain defined by the kernel  $Pf(X)$ , moreover  $(X(k), X'(k))$  satisfy property **P**. In practice, we will start with  $X_0 = X$  and  $X'_0 = X'$ .

We can prove condition (3.6) by the coupon collector's problem.

For this chain, we can write

$$\begin{aligned} m(\theta)' &= \mathbb{E} \operatorname{tr} (f(X) e^{\theta f(X)}) = \mathbb{E} \operatorname{tr} (F(X, X') \cdot e^{\theta f(X)}) \\ &= \frac{1}{2} \mathbb{E} \operatorname{tr} \left( \sum_{k=0}^{\infty} (f(X(k)) - f(X'(k))) \cdot (e^{\theta f(X)} - e^{\theta f(X')}) \right) \\ &= \frac{1}{2} \sum_{k=0}^{\infty} \mathbb{E} \operatorname{tr} \left( \mathbb{1}[I \notin I_1, \dots, I_k] (f(X(k)) - f(X'(k))) \cdot (e^{\theta f(X)} - e^{\theta f(X')}) \right) \end{aligned}$$

Now, using Theorem 2.2, and the fact that  $(f(X(k)) - f(X'(k)))^2 \preceq A_I^2$  and  $(f(X) - f(X'))^2 \preceq A_I^2$ ,

$$\begin{aligned} m(\theta)' &\leq \frac{1}{2} \sum_{k=0}^{\infty} \mathbb{E} \operatorname{tr} \left( \mathbb{1}[I \notin I_1, \dots, I_k] A_I^2 \left( \frac{e^{\theta f(X)} + e^{\theta f(X')}}{2} \right) \right) \\ &\leq \frac{1}{2} \sum_{k=0}^{\infty} \frac{1}{n} \left( 1 - \frac{1}{n} \right)^k \left\| \sum_{i=1}^n A_i^2 \right\| \theta m(\theta) \\ &\leq \frac{1}{2} \sigma^2 \theta m(\theta), \end{aligned}$$

so

$$\log(m(\theta)) \leq \frac{1}{4} \theta^2 \sigma^2,$$

thus by Proposition 3.1,

$$\mathbb{P}\{\lambda_{\max}(\mathbf{H}(Z)) \geq t\} \leq d \cdot \inf_{\theta > 0} \exp \left\{ -\theta t + \frac{1}{4} \theta^2 \sigma^2 \right\} \leq d \cdot \exp \left( -\frac{t^2}{\sigma^2} \right).$$

□

Now we prove the general case:

*Proof for Dobrushin condition.* Let  $X, X', X(k), X'(k)$  be defined analogously to the way it is done in the proof of Theorem 4.3 of Chatterjee (2005):  $X'$  is defined by choosing  $I$  uniformly in  $[n]$ , and then resampling  $X_I$  conditioned on the rest (Gibbs sampler), while  $X(k), X'(k)$  are defined by choosing  $I_k$  uniformly in  $[n]$ , resampling  $X_{I_k}(k-1)$  and  $X'_{I_k}(k-1)$  in the greedy coupling way, i.e.  $X_{I_k}(k)$  is resampled conditionally on the rest of  $X(k-1)$ ,  $X'_{I_k}(k)$  is resampled conditionally on the rest of  $X'(k-1)$ , and at the same time, these two conditional distributions are coupled in the maximal coupling (see Lindvall (1992)). Property **P** can be proven by induction, verifying (3.6) is left to the reader as exercise.

We can write  $f(X(k)) - f(X'(k))$  as a telescopic sum:

$$\begin{aligned} f(X(k)) - f(X'(k)) &= \sum_{i=1}^n f(X_1(k), \dots, X_i(k), X'_{i+1}(k), \dots, X'_n(k)) \\ &\quad - f(X_1(k), \dots, X_{i-1}(k), X'_i(k), \dots, X'_n(k)) =: \sum_{i=1}^n Z_i(k), \end{aligned}$$

We have

$$\begin{aligned}
m'(\theta) &= \mathbb{E} \operatorname{tr} (f(X) e^{\theta f(X)}) = \mathbb{E} \operatorname{tr} (F(X, X') e^{\theta f(X)}) \\
&= \frac{1}{2} \mathbb{E} \operatorname{tr} \left( F(X, X') \left( e^{\theta f(X)} - e^{\theta f(X')} \right) \right) \\
&= \frac{1}{2} \sum_{k=0}^{\infty} \mathbb{E} \operatorname{tr} \left( (f(X(k)) - f(X'(k))) \left( e^{\theta f(X)} - e^{\theta f(X')} \right) \right) \\
&= \frac{1}{2} \sum_{k=0}^{\infty} \sum_{i=1}^n \mathbb{E} \operatorname{tr} \left( Z_i(k) \left( e^{\theta f(X)} - e^{\theta f(X')} \right) \right),
\end{aligned}$$

and obviously  $Z_i(k) = L_i(k)Z_i(k)$ , so by Theorem 2.2, we have

$$\begin{aligned}
&\frac{1}{2} \mathbb{E} \operatorname{tr} \left( L_i(k) Z_i(k) \cdot \left( e^{\theta f(X)} - e^{\theta f(X')} \right) \right) \\
&\leq \frac{1}{2} \mathbb{E} \operatorname{tr} \left( L_i(k) \theta \frac{1}{4} \left( (Z_i(k))^2 + (f(X) - f(X'))^2 \right) \cdot \left( e^{\theta f(X)} + e^{\theta f(X')} \right) \right) \\
&\leq \frac{1}{8} \mathbb{E} \operatorname{tr} \left( L_i(k) \theta (A_I^2 + A_i^2) \cdot \left( e^{\theta f(X)} + e^{\theta f(X')} \right) \right) \\
&\leq \frac{1}{4} \mathbb{E} \operatorname{tr} (l_i(k) \theta (A_I^2 + A_i^2) \cdot e^{\theta f(X)}).
\end{aligned}$$

Let  $D$  be the Dobrushin dependence matrix of  $X_1, \dots, X_n$ , let us denote  $B := (1 - \frac{1}{n}) E + \frac{1}{n} D$ , with  $E$  being the  $n \times n$  identity matrix. Let  $L_i(k) := \mathbb{1}[X_i(k) \neq X'_i(k)]$ , and let  $l_i(k) := \mathbb{E}(L_i(k)|X, X')$ . Page 77-78 of Chatterjee (2005) proves that  $l(k) \leq B^k e(I)$ , with  $e(I)$  denoting the vector whose  $I$ th coordinate is 1 and the rest is 0.

$$\begin{aligned}
&\sum_{i=1}^n \frac{1}{4} \mathbb{E} \operatorname{tr} (l_i(k) (A_I^2 + A_i^2) \cdot \theta e^{\theta f(X)}) \\
&\leq \sum_{i=1}^n \frac{1}{4} \mathbb{E} \operatorname{tr} ([B^k e(I)]_i (A_I^2 + A_i^2) \cdot \theta e^{\theta f(X)}) \\
&= \sum_{i=1}^n \sum_{j=1}^n \frac{1}{n} \frac{1}{4} \mathbb{E} \operatorname{tr} ([B^k e(j)]_i (A_j^2 + A_i^2) \cdot \theta e^{\theta f(X)}) \\
&= \sum_{j=1}^n \frac{1}{n} \frac{1}{4} \mathbb{E} \operatorname{tr} \left( \sum_{i=1}^n [B^k e(j)]_i A_j^2 \cdot \theta e^{\theta f(X)} \right) \\
&\quad + \sum_{i=1}^n \frac{1}{n} \frac{1}{4} \mathbb{E} \operatorname{tr} \left( \sum_{j=1}^n [B^k e(j)]_i A_i^2 \cdot \theta e^{\theta f(X)} \right)
\end{aligned}$$



Now

$$\sum_{i=1}^n [B^k e(j)]_i = \|B^k e(j)\|_1 \leq (\|B\|_1)^k, \text{ and}$$

$$\sum_{j=1}^n [B^k e(j)]_i \leq (\|B\|_\infty)^k,$$

so

$$\frac{1}{2} \mathbb{E} \operatorname{tr} \left( L_i(k) Z_i(k) \cdot \left( e^{\theta f(X)} - e^{\theta f(X')} \right) \right) \leq \frac{1}{4n} \sigma^2 \theta m(\theta) \left( (\|B\|_1)^k + (\|B\|_\infty)^k \right).$$

Summing up in  $k$ , and noticing that  $\|B\|_1 \leq 1 - \frac{1}{n} + \frac{1}{n} \|D\|_1$  gives

$$m'(\theta) \leq \frac{\theta}{4} \sigma^2 m(\theta) \left( \frac{1}{1 - \|D\|_1} + \frac{1}{1 - \|D\|_\infty} \right),$$

and thus we get the result by Proposition 3.1, as in the independent case.  $\square$

$\square$

#### 4. Proof of trace inequalities

Before starting the proof, we state a few simple inequalities:

- For any  $P, Q \in \mathbb{M}^d$ , we have

$$PQ + Q^* P^* \preceq PP^* + Q^* Q, \quad (4.1)$$

which follows from  $(P + Q^*)(P^* + Q) \succeq 0$ .

- Also, we can easily prove that if  $P, Q, R, S \in \mathbb{H}^d$ , then

$$\operatorname{Re}(\operatorname{tr}(PQRS)) \leq \operatorname{tr} \left( \frac{(P^2 + R^2)(Q^2 + S^2)}{4} \right) \quad (4.2)$$

To prove this, just apply (4.1) to  $(PQ)(RS)$  and to  $(QR)(SP)$ , and rearrange the terms.

*Proof of Theorem 2.2.* First notice that adding a constant times identity matrix to  $A$  and  $B$  multiplies both sides by the same number. Therefore we can suppose without loss of generality that  $A, B \succeq 0$ .

By taking Taylor expansion, the inequality becomes

$$\operatorname{tr} \left( C \sum_{k=1}^{\infty} \frac{A^k - B^k}{k!} \right) \leq \operatorname{tr} \left( \frac{C^2 + (A - B)^2}{4} \sum_{k=1}^{\infty} \frac{A^{k-1} + B^{k-1}}{(k-1)!} \right)$$

To show this, we will prove that the inequality holds for each term in the sums, i.e. we claim:

$$\operatorname{tr} (C(A^k - B^k)) \leq k \cdot \operatorname{tr} \left( \frac{C^2 + (A - B)^2}{4} (A^{k-1} + B^{k-1}) \right) \quad (4.3)$$

Now

$$A^k - B^k = A^k - A^{k-1}B + A^{k-1}B - A^{k-2}B^2 + \dots + AB^{k-1} - B^k.$$

The terms in the sum are of the form  $A^l B^{k-l} - A^{l-1} B^{k-l+1} = A^{l-1}(A - B)B^{k-l}$ . We claim the following about two 'symmetric' terms from such a sum (which clearly implies (4.3)):

**Lemma 4.1.** *If  $A, B, C$  are Hermitian matrices with  $A, B$  positive definite, and  $0 \leq k \leq n$  are integers, then*

$$\begin{aligned} & \operatorname{Re} \left( \operatorname{tr} \left( C \left( A^k (A - B) B^{n-k} + A^{n-k} (A - B) B^k \right) \right) \right) \\ & \leq \operatorname{tr} \left( \frac{C^2 + (A - B)^2}{2} (A^n + B^n) \right). \end{aligned} \quad (4.4)$$

*Proof.* Let us denote  $D := A - B$ , then the inequality becomes

$$\operatorname{Re} \left( \operatorname{tr} \left( C \left( A^k D B^{n-k} + A^{n-k} D B^k \right) \right) \right) \leq \operatorname{tr} \left( \frac{C^2 + D^2}{2} (A^n + B^n) \right) \quad (4.5)$$

If  $k = n/2$ , then this follows from (4.2). Suppose, without loss of generality, that  $k < n/2$ . Now we can get rid of  $C$  in the following way:

$$\begin{aligned} & \operatorname{Re} \left( \operatorname{tr} \left( C A^k D B^{n-k} + C A^{n-k} D B^k \right) \right) = \\ & \operatorname{Re} \left( \operatorname{tr} \left( (B^{n/2} C) (A^k D B^{n/2-k}) + (C A^{n/2}) (A^{n/2-k} D B^k) \right) \right) \leq \\ & \leq \frac{1}{2} \operatorname{Re} \left( \operatorname{tr} \left( C^2 B^n + A^{2k} D B^{n-2k} D + C^2 A^n + A^{n-2k} D B^{2k} D \right) \right) \end{aligned}$$

The terms involving  $C$  are the same as on the right hand side of (4.4), so we need to prove that

$$\operatorname{tr} \left( A^{2k} D B^{n-2k} D + A^{n-2k} D B^{2k} D \right) \leq \operatorname{tr} \left( D^2 (A^n + B^n) \right). \quad (4.6)$$

Both sides are real so we did not write the real part.

Let us denote, for  $0 \leq l \leq n$ ,

$$\begin{aligned} R &:= \operatorname{tr} \left( D^2 (A^n + B^n) \right), \text{ and} \\ T_l &:= \operatorname{tr} \left( A^l D B^{n-l} D + A^{n-l} D B^l D \right). \end{aligned}$$

We can show that when  $l = n/2$ , then  $T_l \leq R$  holds. Let us denote the maximum of  $T_l$  as  $T := \max_{1 \leq l \leq n} T_l$ , and suppose that the maximum is taken at  $l_0 < n/2$ , then

$$\begin{aligned} & \operatorname{tr} \left( A^{l_0} D B^{n-l_0} D + A^{n-l_0} D B^{l_0} D \right) \\ & = \operatorname{tr} \left( (A^{l_0} D B^{n/2-l_0}) (B^{n/2} D) + (A^{n/2} D) (B^{l_0} D A^{n/2-l_0}) \right), \end{aligned}$$

so as previously, using (4.1) we can show that  $T_{l_0} \leq \frac{1}{2} T_{2l_0} + \frac{1}{2} R$ , i.e.  $T \leq \frac{1}{2} T + \frac{1}{2} R$ , and thus  $T \leq R$ .

For  $l_0 > n/2$ , we can show that  $T_{l_0} \leq \frac{1}{2} T_{2n-2l_0} + \frac{1}{2} R$ . □

□

*Proof of Theorem 2.3.* (2.7) is a generalization of (4.5). If  $p$  is rational of the form  $\frac{a}{b}$ , then we can proceed the same way as in the proof of Lemma 4.1, with  $A^{\frac{1}{b}}$  and  $B^{\frac{1}{b}}$  taking the role of  $A$  and  $B$ . If  $p$  is irrational, we can write it as the limit of rationals, and use continuity to get the result. □

## 5. Open problems

Based on extensive numerical evidence and on some theoretical results, we conjecture the following trace inequalities:

1. Let  $A, B, C \in \mathbb{H}^d$ , then

$$\begin{aligned} & \operatorname{tr}(C(e^A - e^B)) \\ & \leq \operatorname{tr}\left(\frac{C_+^2 + (A - B)_+^2}{2}e^A + \frac{C_-^2 + (A - B)_-^2}{2}e^B\right) \end{aligned} \quad (5.1)$$

2. We expect (5.1) to generalize to any monotone increasing convex function  $f$ , i.e. we expect that in such situations,

$$\begin{aligned} & \operatorname{tr}(C(f(A) - f(B))) \\ & \leq \operatorname{tr}\left(\frac{C_+^2 + (A - B)_+^2}{2}f'(A) + \frac{C_-^2 + (A - B)_-^2}{2}f'(B)\right) \end{aligned} \quad (5.2)$$

3. (5.1) would imply concentration for self-bounding matrix valued functions (in the sense of [Boucheron, Lugosi and Massart \(2009\)](#)). A similar setting has been already studied in [Mackey \(2012\)](#), Theorem 25, however, this theorem requires a very strong self-reciprocity condition, which may not be satisfied in general.

We define matrix self-bounding functions as follows:

**Definition 2.** An  $\mathbb{H}^d$  valued function  $\mathbf{H}(Z_1, \dots, Z_n)$  is said to be  $(a, b)$  matrix self-bounding, if for any  $Z'_1, \dots, Z'_n$ ,

$$(a) \quad \mathbf{H}(Z) - \mathbf{H}(Z_1, \dots, Z'_i, \dots, Z_n) \leq \mathbf{I}_d, \text{ and}$$

$$(b) \quad \sum_{i=1}^n (\mathbf{H}(Z) - \mathbf{H}(Z_1, \dots, Z'_i, \dots, Z_n))_+ \preceq a\mathbf{H}(Z) + b\mathbf{I}_d.$$

An  $\mathbb{H}^d$  valued function  $\mathbf{H}(Z_1, \dots, Z_n)$  is said to be weakly  $(a, b)$  matrix self-bounding, if for any  $Z'_1, \dots, Z'_n$ ,

$$\sum_{i=1}^n (\mathbf{H}(Z) - \mathbf{H}(Z_1, \dots, Z'_i, \dots, Z_n))_+^2 \preceq a\mathbf{H}(Z) + b\mathbf{I}_d.$$

We expect concentration inequalities similar to Theorem 1 of [Boucheron, Lugosi and Massart \(2009\)](#) to hold for such functions.

## Acknowledgements

The author thanks Doma Szász and Mogyi Tóth for infecting him with their enthusiasm of probability. He thanks his thesis supervisors, Louis Chen and Adrian Röllin, for the opportunity to study in Singapore, and their useful advices. Finally, many thanks to my brother, Roland Paulin, for the enlightening discussions.

## References

- AHLSWEDE, R. and WINTER, A. (2002). Strong converse for identification via quantum channels. *IEEE Trans. Inform. Theory* **48** 569–579. . [MR1889969 \(2003d:94069\)](#)
- BOUCHERON, S., LUGOSI, G. and MASSART, P. (2009). On concentration of self-bounding functions. *Electron. J. Probab.* **14** no. 64, 1884–1899. [MR2540852 \(2010k:60058\)](#)
- CHATTERJEE, S. (2005). *Concentration inequalities with exchangeable pairs*. Thesis (Ph.D.)–Stanford University, Available at <http://arxiv.org/abs/math.PR/0507526>. [MR2707160](#)
- LINDVALL, T. (1992). *Lectures on the coupling method*. *Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics*. John Wiley & Sons Inc., New York. A Wiley-Interscience Publication. [MR1180522 \(94c:60002\)](#)
- MACKEY, L. (2012). Matrix Factorization and Matrix Concentration PhD thesis, EECS Department, University of California, Berkeley Available at <http://www.eecs.berkeley.edu/Pubs/TechRpts/2012/EECS-2012-99.html>.
- MACKEY, L., JORDAN, M. I., CHEN, R. Y., FARRELL, B. and TROPP, J. A. (2012). Matrix Concentration Inequalities via the Method of Exchangeable Pairs. *ArXiv e-prints*.
- PETZ, D. (1994). A survey of certain trace inequalities. In *Functional analysis and operator theory (Warsaw, 1992)*. *Banach Center Publ.* **30** 287–298. Polish Acad. Sci., Warsaw. [MR1285615 \(95c:15038\)](#)
- TROPP, J. A. (2012). User-friendly tail bounds for sums of random matrices. *Found. Comput. Math.* **12** 389–434. . [MR2946459](#)